VISIT…

Nucleotides and Nucleic Acids

CHAPTER 27

MASTERING ORGANIC CHEMISTRY

- Understanding the General Structures of Nucleosides, Nucleotides, DNA, and RNA
- Recognizing the Hydrogen Bonding That Occurs between Complementary Bases
- Understanding the General Features of Replication, Transcription, and Translation
- Understanding How the Sequence of DNA Is Determined
- Understanding How DNA Is Synthesized in the Laboratory

THE NUCLEIC ACIDS, DNA and RNA, store genetic information in living organisms and are responsible for translating this information into the structure of proteins. Because their structure and function are discussed in great detail in modern biochemistry texts, this chapter concentrates on the organic chemical aspects of these important biomolecules.

The chapter begins with a discussion of the structure of nucleosides and nucleotides. Then the structure of the nucleic acids, DNA and RNA, the polymers formed from nucleotide monomers, is presented. The function of these polymers in the replication, transcription, and translation of genetic information is briefly addressed. Next, the organic chemistry involved in determining the sequence of DNA is presented. Finally, the synthesis of small DNA molecules in the laboratory is discussed.

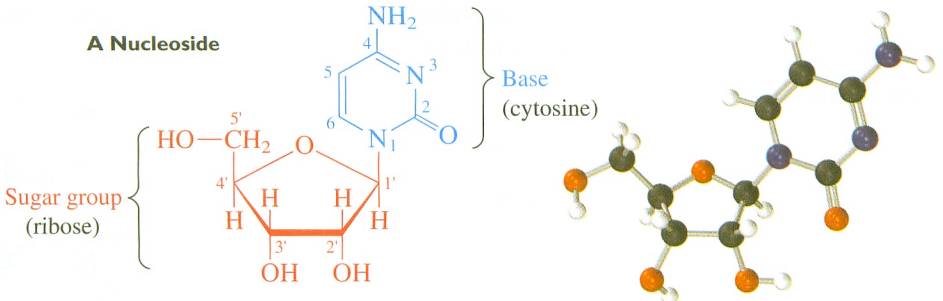
27.1 NUCLEOSIDES AND NUCLEOTIDES

Nucleosides are composed of a sugar, either ribose or 2′-deoxyribose, and a base, a cyclic nitrogen-containing compound. These groups are connected by a bond between one of the nitrogen atoms of the base and the anomeric carbon (acetal carbon) of the sugar (a nitrogen analog of a glycoside bond). The nucleoside cytidine is formed from the sugar ribose and the base cytosine:

ORGANIC Chemistry Now™

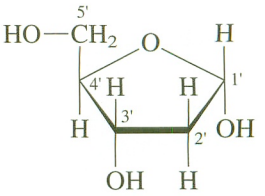
Look for this logo in the chapter and go to OrganicChemistryNow at **http://now.brookscole.com/hornback2** for tutorials, simulations, problems, and molecular models.

A Nucleoside



Cytidine

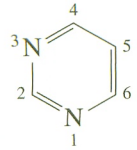
In nucleosides the atoms of the heterocyclic base are numbered 1, 2, and so on, and the positions of the sugar are numbered with primes, that is, 1′, 2′, and so on. Based on this numbering system, the base is attached to the 1′ position of the sugar. Nucleosides may also be formed from the sugar 2′-deoxyribose, which differs from ribose only in that the hydroxy group at the 2′ position has been replaced with a hydrogen.



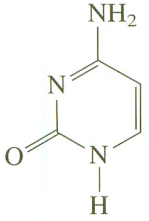
2'-Deoxyribose

Five nitrogen bases are commonly found in nucleosides. Three of these—cytosine, thymine, and uracil—are based on the pyrimidine ring system. They are attached to the sugar by the nitrogen at position 1. The other two bases—adenine and guanine—are based on the purine ring system. They are attached to the sugar by the nitrogen at position 9:

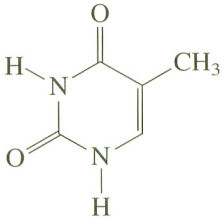
Nitrogen bases



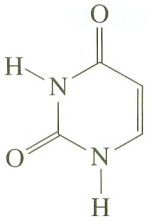
Pyrimidine



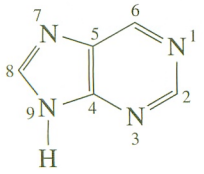
Cytosine



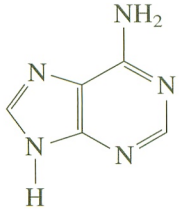
Thymine



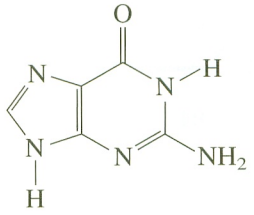
Uracil



Purine



Adenine



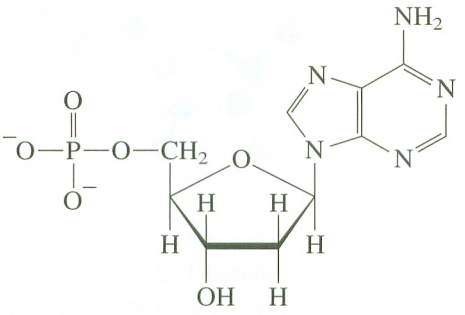
Guanine

The names of the nucleosides formed from these bases and ribose are cytidine, thymidine, uridine, adenosine, and guanosine. Nucleosides formed from deoxyribose are named using the prefix deoxy-, as in deoxythymidine.

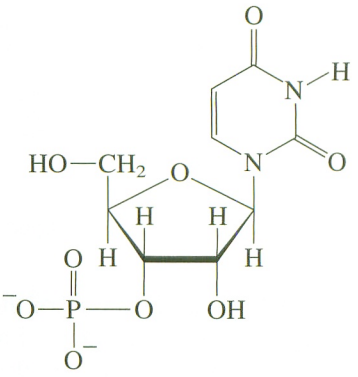
Nucleotides differ from nucleosides in that they have phosphate groups attached to either the 3′ or 5′ position of the sugar. The structures of deoxyadenosine 5′-monophosphate,

a nucleotide formed from adenine and deoxyribose, and uridine 3′-monophosphate, a nucleotide formed from uracil and ribose, are as follows:

Nucleotides



Deoxyadenosine 5'-monophosphate



Uridine 3'-monophosphate

PROBLEM 27.1

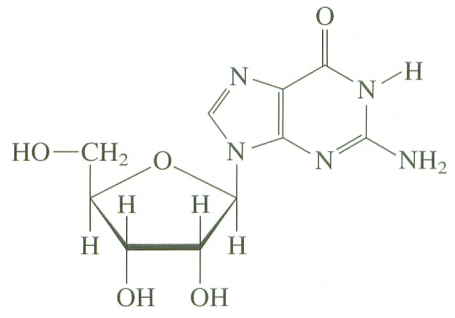
Draw the structures of these compounds:

a) Thymidine
b) Deoxyguanosine
c) Deoxyuridine 5′-monophosphate
d) Deoxycytidine 3′-monophosphate

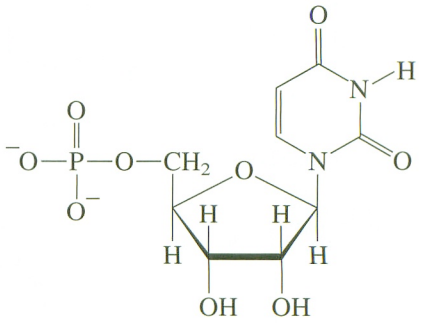
PROBLEM 27.2

Provide names for these compounds:

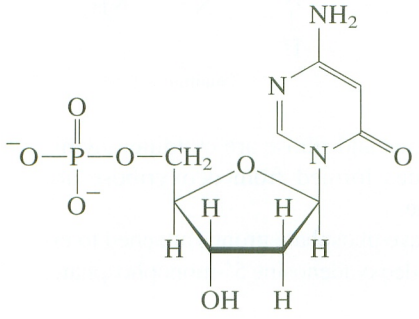
a)



b)



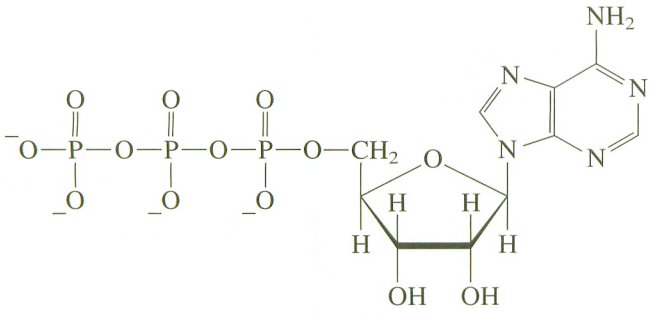
c)



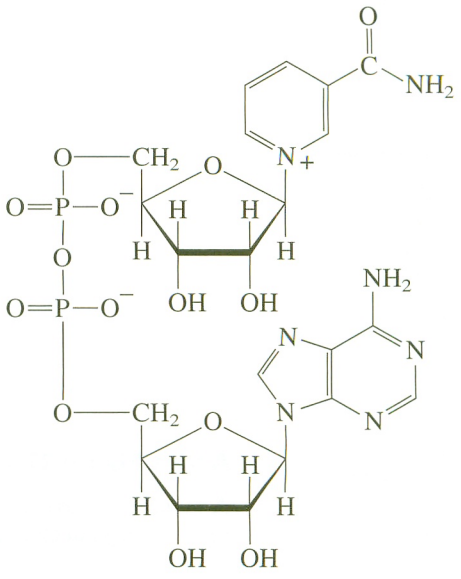
PROBLEM 27.3
Explain why pyrimidine and purine are aromatic compounds.

PROBLEM 27.4
Show a resonance structure for each of these compounds that emphasizes its aromatic character:
a) Guanine
b) Cytosine
c) Thymine

In addition to their role in the formation of DNA and RNA (see Section 27.2), nucleotides have other important biological functions. For example, adenosine triphosphate (ATP) is an important energy carrier in biochemical reactions, and nicotinamide adenine dinucleotide is a coenzyme that is often involved in biochemical oxidation–reduction reactions.



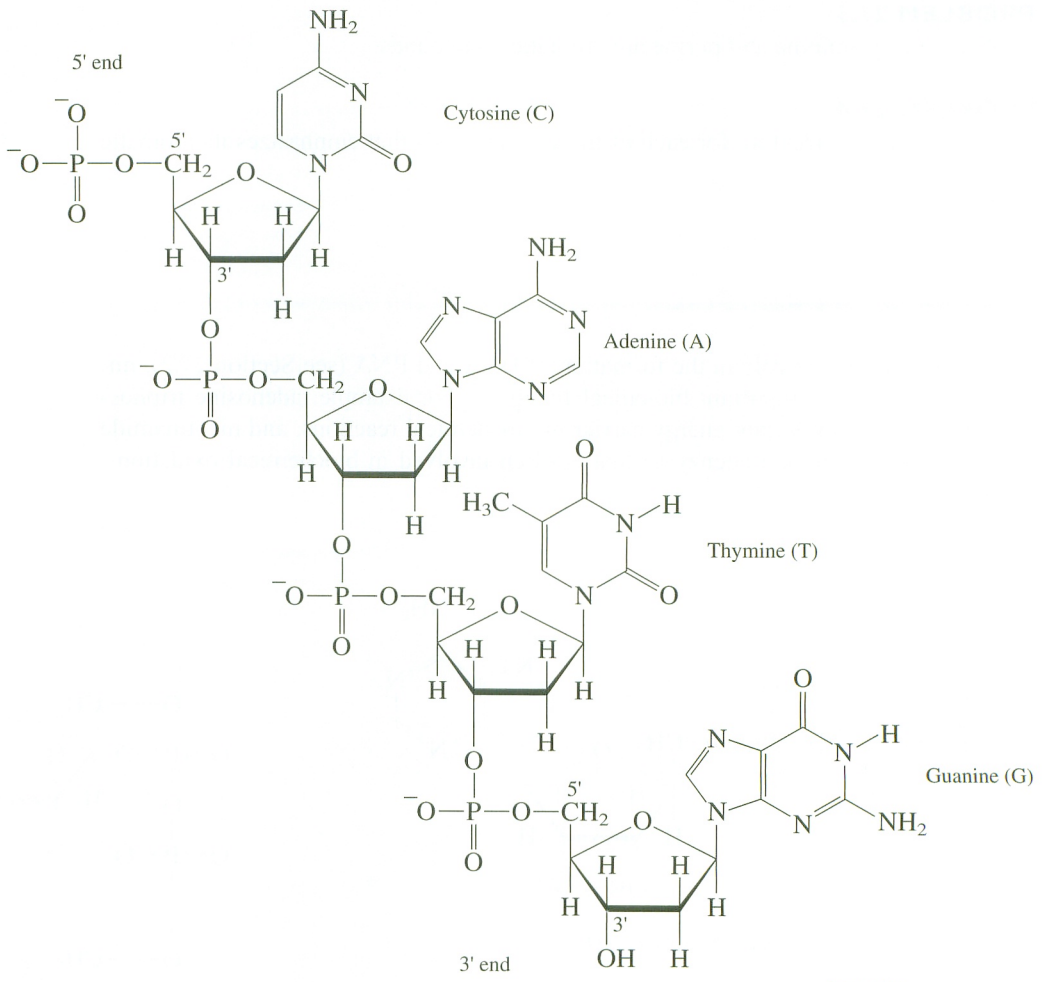
Adenosine triphosphate (ATP)



Nicotinamide adenine dinucleotide (NAD^+)

27.2 STRUCTURE OF DNA AND RNA

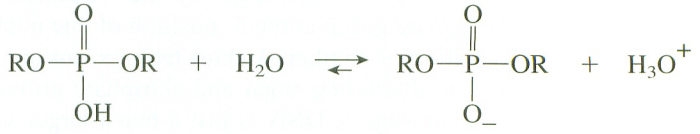
Nucleotides have several functional groups, so they can form condensation-type polymers. The formation of deoxyribonucleic acid (DNA) can be viewed as resulting from the polymerization of deoxyribonucleotides by the formation of a phosphate ester bond between a phosphate group at the 5′ position of one nucleotide and the hydroxy group at the 3′ position of another. A short tetranucleotide is shown in Figure 27.1. It has a backbone of alternating sugar and phosphate groups, with the heterocyclic bases appended to the sugars. DNA is just a much larger version of a



Active Figure 27.1 ORGANIC Chemistry Now™

A TETRANUCLEOTIDE WITH THE GENERAL STRUCTURE OF DNA. Test yourself on the concepts in this figure at **OrganicChemistryNow.**

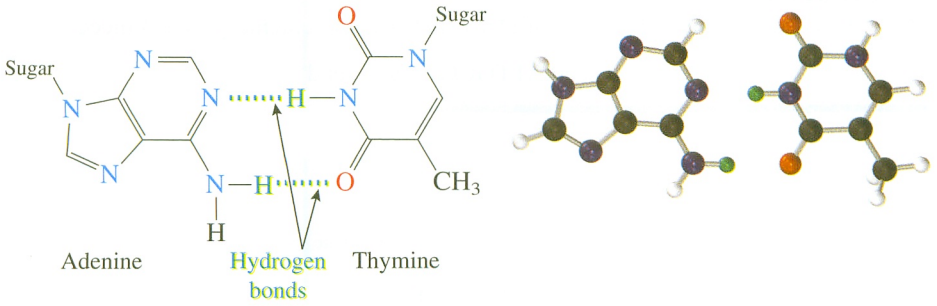
polymer with this general structure. If the negative oxygens of the phosphate groups of DNA were protonated, then each of these groups would be a diester of phosphoric acid. As shown in the following equation, this ester is a relatively strong acid, comparable in strength to phosphoric acid, which has $pK_{a1} = 2.2$. Therefore, nucleic acids are strong acids and are completely ionized at the near neutral pH that occurs in living organisms.



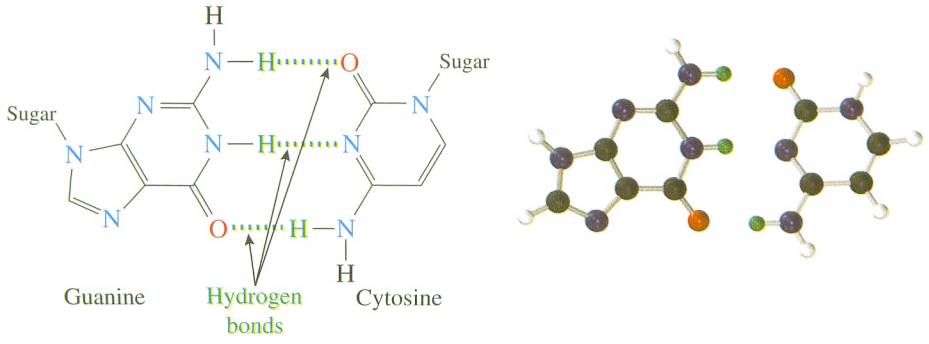
By the early 1950s, DNA was known to have the general structure shown in Figure 27.1. However, how these molecules are arranged in the cell, how they are used to store genetic information, and how they provide a mechanism for the replication of that genetic information were a mystery until James Watson and Francis Crick solved the puzzle. In 1953, Watson and Crick published a paper that proposed an explanation for the overall structure of DNA and showed that certain features of this structure explained how genetic information was replicated when the cell divided. This paper is certainly one of the most important scientific publications ever and is generally recognized to mark the beginning of modern biochemistry and molecular biology. Watson, Crick, and M. Wilkins were awarded the 1962 Nobel Prize in physiology or medicine for this work. Wilkins obtained the X-ray diffraction patterns that provided some vital information about the structure of DNA.

A number of clues enabled Watson and Crick to make this important discovery. Wilkins's X-ray photographs suggested to Crick, an X-ray crystallographer by background, that DNA was a helical molecule with its aromatic bases arranged in a planar stack. Another clue was found in Chargaff's rules, which state that the number of adenine residues in a particular DNA molecule is always equal to the number of thymine residues and that the number of guanine residues is always equal to the number of cytosine residues. Using this information and building numerous models enabled Watson and Crick to solve the DNA puzzle.

The explanation for Chargaff's rules is provided by the selective hydrogen bonds, termed **complementary base pairing,** that form between adenine and thymine molecules and between guanine and cytosine. Adenine and thymine form two hydrogen bonds with an overall strength of about 10 kcal/mol (42 kJ/mol):



Guanine and cytosine form three hydrogen bonds with an overall strength of about 17 kcal/mol (71 kJ/mol):



ORGANIC Chemistry Now™
Click *Coached Tutorial Problems* for more experience with **Complementary Base Pairing.**

These interactions are strong enough to bind one DNA chain to its complementary chain under most circumstances, thus providing stability for the storage of genetic information. However, on occasions when the chains need to be separated—that is, when they are to be replicated or used to delineate the sequence of amino acids in a protein that is being synthesized—the hydrogen bonds are weak enough to be readily broken.

MODEL BUILDING PROBLEM 27.1

Build models to show the hydrogen bonding that occurs in these base pairs:

a) Adenine and thymine

b) Guanine and cytosine

The model proposed by Watson and Crick for DNA has two polymer chains running in opposite directions and held together by hydrogen bonds between complementary bases. A schematic representation of this structure is shown in Figure 27.2. The left chain has its 5′ end at the top of the page and its 3′ end at the bottom. The right chain has its 5′ end at the bottom of the page and its 3′ end at the top. Note that the bases in the chains are not identical. Instead, they are complementary; that is, wherever one chain has an adenine, the other has a thymine and so forth. Because of the specific hydrogen bonds that form between complementary base pairs, the order of the bases in one chain dictates the order in the other. This means that both chains provide enough information to allow the construction of the other chain. This is used in replication and also helps to ensure that the information in the base sequence is maintained with high fidelity.

PROBLEM 27.5

Show the base sequence of a piece of DNA that is complementary to this piece:

3′ end ATTGCGAGC 5′ end

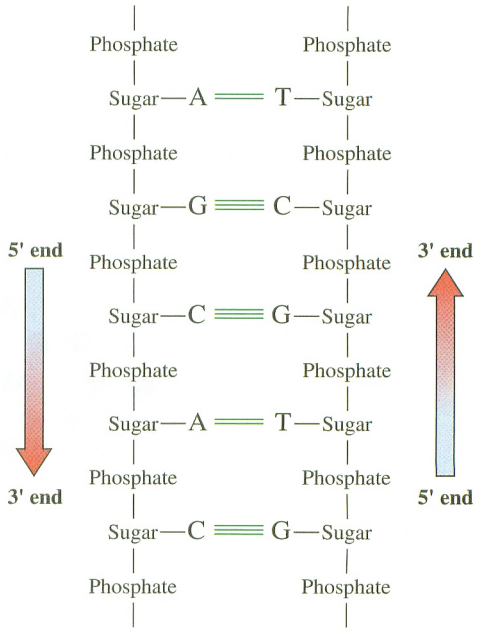


Figure 27.2

SCHEMATIC REPRESENTATION OF BASE PAIRING BETWEEN DNA CHAINS. A = adenine, T = thymine, G = guanine, C = cytosine.

As shown in Figure 27.3, the two chains of this dimeric polymer coil around each other in a helical manner, producing the famous DNA double helix. The distance between the sugar–phosphate backbones of the two chains remains constant, regardless of which bases connect them because a purine base (adenine or guanine) is always hydrogen bonded to a pyrimidine base (thymine or cytosine). The less polar bases are on the inside of the helix, where their interaction with the water molecules of the aqueous environment is minimized, whereas the ionic phosphate groups are on the outer surface of the helix, where they can be readily solvated. In addition, the planes of the bases are stacked one on top of another. The interactions between these stacked bases are important in stabilizing the double helix.

ORGANIC Chemistry Now™
Click *Coached Tutorial Problems* to view more on **DNA Structure.**

DNA is a truly enormous molecule. Human DNA contains approximately 3 billion base pairs grouped into 23 individual DNA molecules or chromosomes. Each base pair contributes 3.4 Å to the length of a DNA molecule. Therefore, the total length of the DNA in the 23 human chromosomes is approximately $(3.4 \times 10^{-10}\ \text{m})(3 \times 10^{9}) \cong 1\ \text{m}$, but it is only 20 Å in diameter. Obviously, DNA is highly coiled in the cell.

The chemical behavior of DNA is primarily what would be expected on the basis of its structure. The sugar–phosphate backbone is held together by phosphate ester bonds. These bonds are not too different from carboxylic ester bonds. Thus, the phosphodiester

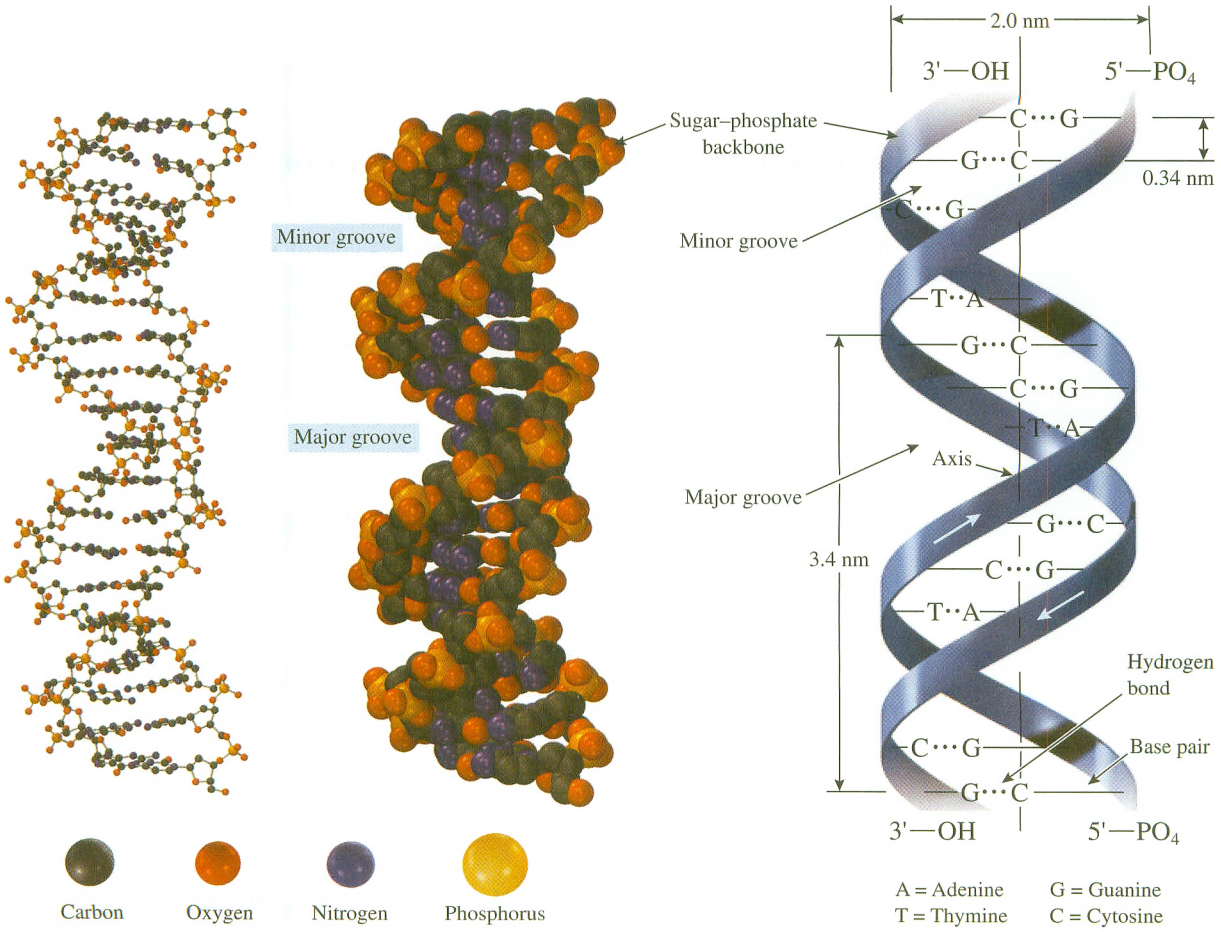


Figure 27.3

MODEL OF THE DNA DOUBLE HELIX. Hydrogens have been omitted for clarity.

linkages can be hydrolyzed in acidic solution. The heterocyclic bases are attached to the ribose or deoxyribose groups by a nitrogen analog of a glycoside bond. Like its oxygen analog, the N-glycosidic bond is also hydrolyzed by aqueous acid.

RNA has the same general structure as DNA with several exceptions. It employs the sugar ribose rather than deoxyribose. It has the base uracil in place of thymine. Note that thymine and uracil are both pyrimidine bases, and both hydrogen bond to adenine in the same manner. Unlike DNA, RNA usually occurs as a single-stranded molecule. Because RNA is used to transfer and translate the information that is stored in DNA the complementary copy is not necessary and is not synthesized by the cell. Finally, RNA is much smaller than DNA.

PROBLEM 27.6

Use drawings to show the hydrogen bonding that occurs in a base pair formed from adenine and uracil.

PROBLEM 27.7

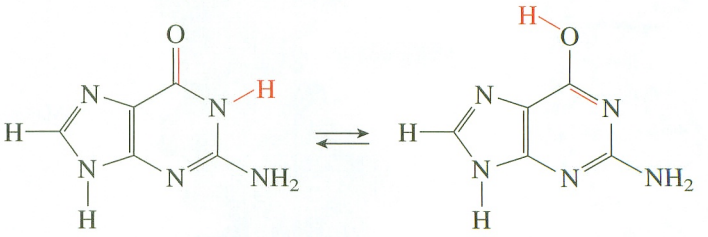
Use drawings to explain why thymine and guanine do not form a base pair that has hydrogen bonds as strong as those between cytosine and guanine.

Focus On

Tautomers of Guanine and Thymine

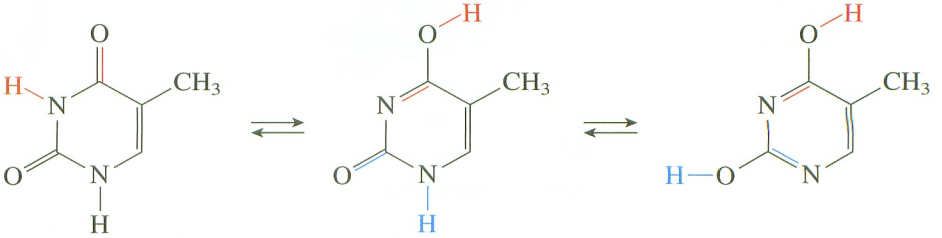
The specific hydrogen bonding that occurs between adenine and thymine or guanine and cytosine was of crucial importance in the development of the model for the structure of DNA by Watson and Crick. However, at that time, many people believed that guanine existed primarily as the enol tautomer and thymine as the dienol tautomer because both of these structures would be fully aromatic.

Guanine



Keto tautomer Enol tautomer

Thymine



Keto tautomer Enol tautomer Dienol tautomer

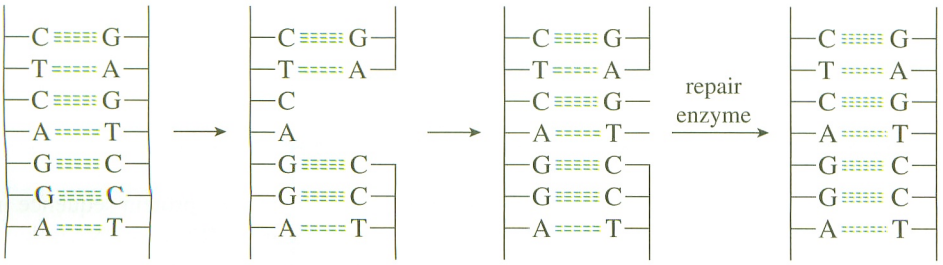
Actually, because of the strength of the CO double bond and the fact that not much resonance energy is lost, the keto tautomers are more stable. Fortunately, Watson's officemate was an expert on X-ray structures of small organic molecules and knew that the keto isomers were the predominant tautomers. Without this information, Watson and Crick would probably have used the incorrect structures in their model-building studies. Of course, the enol tautomers do not form the same specific hydrogen bonding interactions with the other bases, so this important clue to the structure and function of DNA would have remained hidden. Luck sometimes plays an important role in great discoveries.

27.3 Replication, Transcription, and Translation

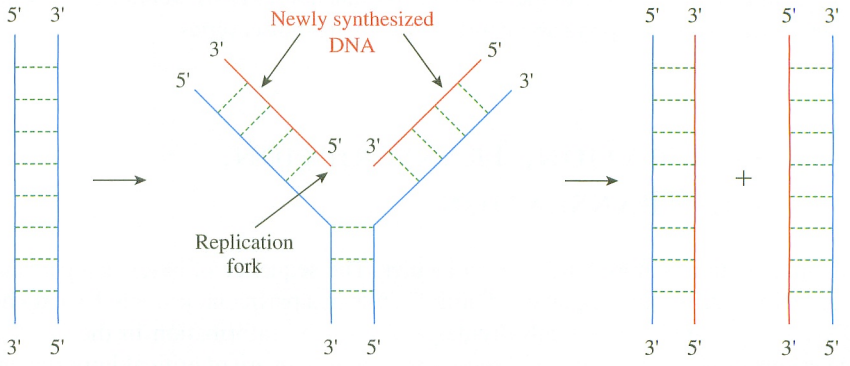
Genetic information is stored in DNA molecules. The sequence of bases in a particular piece of DNA specifies the sequence of amino acids in a particular protein. Exactly how the DNA is replicated when a cell divides and how the information in the DNA sequence is converted to an amino acid sequence of a protein are of critical importance to the functioning of a living organism and comprise a major part of any modern biochemistry textbook. We have space here only to briefly outline these processes.

The information for the amino acid sequence of a protein is stored in the base sequence of DNA. However, 20 amino acids are found in proteins, whereas only 4 bases occur in DNA. This means that a single base cannot code for an individual amino acid. Likewise, a two-base code, which provides $4 \times 4 = 16$ combinations, still is not large enough to specify 20 different amino acids. The genetic code is actually based on a series of three bases, called a **codon**, which provides $4^3 = 64$ different possibilities. A codon for a particular amino acid is designated by listing the first letters of the three bases that compose it. Thus, one codon for serine is UCA, which designates a base sequence of uracil, cytosine, and adenine. The code is degenerate; that is, most amino acids are specified by two or more codons. For example, the codons CCC, CCU, CCA, and CCG all specify the amino acid proline. In addition, the codons UAA, UAG, and UGA all specify a stop signal; that is, they indicate the end of the protein chain.

The two complementary strands of DNA provide a stable reservoir for genetic information, which must be preserved over the entire lifetime of the organism. Either of the strands, by itself, has all of the genetic information. Therefore, if one strand is damaged, perhaps by random hydrolysis, the information to repair the damage is still present in the complementary strand. For example, suppose a G and a T were lost from the piece of DNA represented by the following schematic structure. When a repair enzyme encounters this damage, the C and A of the complementary strand ensure that the correct bases are inserted in the repair process.



The double-stranded nature of DNA also provides a method for **replication**, the process whereby DNA is duplicated so that two identical copies are available when a cell divides. In this process, the DNA unwinds, and each strand serves as a template for the synthesis of its complementary strand. When replication is completed, two identical versions of the original DNA helix are present. This process is represented schematically in the following diagram:



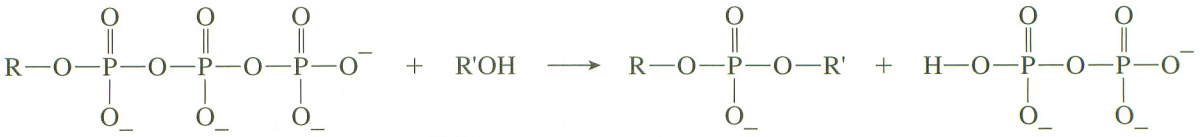
Original DNA Unwinding and synthesis Duplicate copies of original DNA

Like all biological processes, replication is controlled by enzymes. Enzymes unwind the DNA double helix so that the individual chains are accessible. Then an enzyme called DNA polymerase attaches a new nucleotide to a deoxyribose in the growing chain of the new DNA strand.

In replication, the nucleotide that is to be attached next is held in place by the DNA polymerase enzyme. This new nucleotide is complementary to the base in the original DNA strand. The reason usually given for the incorporation of the correct nucleotide (the complementary nucleotide) is the favorable hydrogen bonding that occurs between complementary bases. However, recent studies have shown that the size and shape of the base is at least as important. As shown in Figure 27.4, the new nucleotide has a triphosphate group attached to its 5′-hydroxy group. The 3′-hydroxy group on the end of the growing DNA chain reacts with the triphosphate to form a phosphate ester in a reaction that is quite similar to the reaction of an alcohol with a carboxylic acid anhydride to form the ester of a carboxylic acid. Of course, these reactions are under enzymatic control.

PROBLEM 27.8

The mechanism for the attachment of a new nucleotide to a growing DNA chain shown in Figure 27.4 is very similar to the mechanism for the formation of a carboxylic ester from an alcohol and a carboxylic anhydride. Suggest the steps in the mechanism for this reaction:



The transformation of the information in a DNA strand into a protein sequence involves several steps. First an RNA polymer, called **messenger RNA** (mRNA), that is complementary to the DNA is synthesized in much the same manner as the synthesis of new DNA described earlier. This process is called **transcription.** Messenger RNA

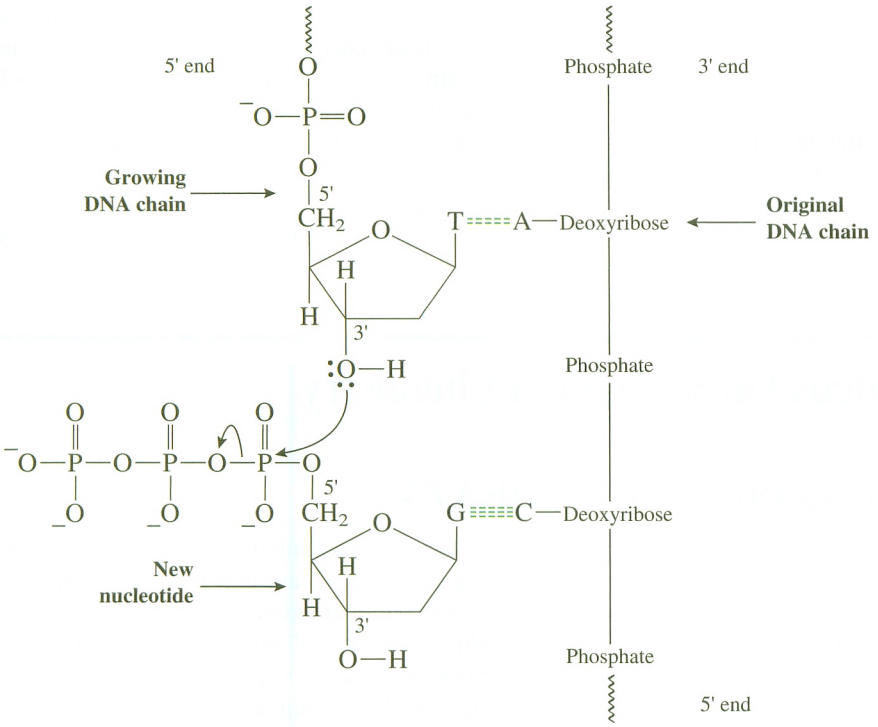
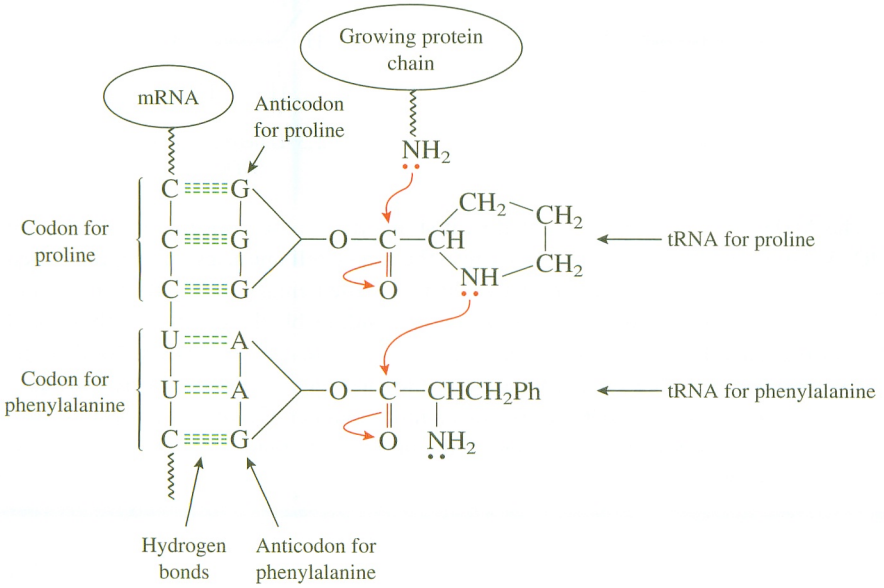


Figure 27.4

ATTACHMENT OF A NEW NUCLEOTIDE TO A GROWING DNA CHAIN.

then serves as a template for protein synthesis in a process called **translation.** Individual amino acids are attached to relatively small RNA molecules called **transfer RNA** (tRNA). Each amino acid has its own type of tRNA that has a three-base region, known as the **anticodon,** that is complementary to the codon for that amino acid. The tRNA with the correct amino acid forms three base pairs with the mRNA and brings the amino acid into position for attachment to the growing protein chain. The process for attaching a proline, followed by a phenylalanine, to a growing polypeptide chain is represented schematically in the following diagram:



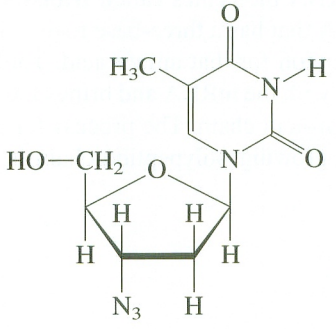
The tRNA for proline, which has the anticodon GGG, hydrogen bonds to the CCC codon for proline in the mRNA and brings its attached proline amino acid into position for attachment to the growing protein chain. Then the tRNA for phenylalanine hydrogen bonds to the codon for phenylalanine in the mRNA and brings its attached phenylalanine into position for attachment to the proline. This process continues until a stop signal is reached.

This has been a necessarily brief overview of the processes of replication, transcription, and translation. If you are interested in learning more, please consult a biochemistry textbook.

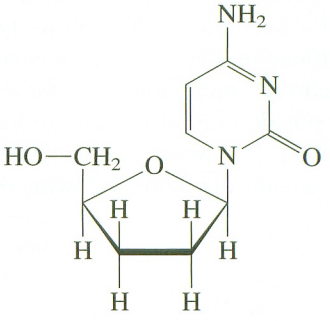
Focus On Biological Chemistry

Treatment of AIDS with AZT

Azidothymidine, also known as AZT or zidovudine, was initially prepared in the 1960s with the hope that it might be useful in the treatment of cancer. It is similar to the nucleoside deoxythymidine, except that it has an azido group in place of the 3′-hydroxy group. The reasoning was that this "fraudulent nucleoside" might be incorporated into new DNA chains that were being synthesized by the cancer cells. Because AZT has no 3′-hydroxy group, it would act to terminate the DNA chain and would therefore prevent cell division. Unfortunately, cancer cells recognize AZT as a fake and do not incorporate it.



Azidothymidine
(AZT)



2',3'-Dideoxycytidine
(DDC)

The AIDS virus is a retrovirus; that is, it stores its genetic information in the form of RNA. On infection it injects its RNA into the target cell and uses an enzyme called reverse transcriptase to synthesize DNA that is complementary to this RNA template. AZT is accepted by reverse transcriptase as a building block for this synthesis and slows or prevents the conversion of the viral RNA information into DNA. By disrupting this process, AZT slows the replication of the virus in the cell. The nucleoside analog 2′,3′-dideoxycytidine (DDC) works in a similar manner. Current treatments, which use a cocktail of two nucleoside analogs and a protease inhibitor, seem to at least slow the progress of the disease.

27.4 SEQUENCING DNA

Determination of the sequence of bases in DNA is accomplished in a manner similar to the determination of the sequence of amino acids in a protein. The problem is somewhat simpler in that there are only four possible bases. However, DNA is much larger than a protein, so there is considerably more information to obtain. First it is necessary to break the enormous DNA molecule into specific fragments of more manageable size. Then the sequence of bases in these fragments must be determined. Several different cleavage processes must be employed to produce the fragments so that one set overlaps the cleavage points of another and can be used to determine how the fragments were connected.

Until recently, sequencing DNA was difficult because no methods were available to cleave the DNA at specific positions. In addition, no reaction analogous to the Edman degradation is available to remove one base at a time for analysis. However, several recent developments have simplified the sequencing of DNA considerably, so it is now easier than the sequencing of a protein. For this reason the sequence of a protein is often determined today by isolating the DNA that codes for that protein and sequencing it. Let's see how DNA sequencing is accomplished.

The discovery of **restriction endonucleases** was of crucial importance in sequencing DNA. These enzymes recognize a specific sequence of four to eight bases in double-stranded DNA and cleave the DNA at a precise point in this sequence. For example, the restriction endonuclease known as *Alu*I cleaves the sequence AGCT between the G and C and the one known as *Pst*I cleaves the sequence CTGCAG between the A and the G. These restriction endonuclease enzymes provide a reproducible way to produce precisely defined fragments of an appropriate size for sequencing.

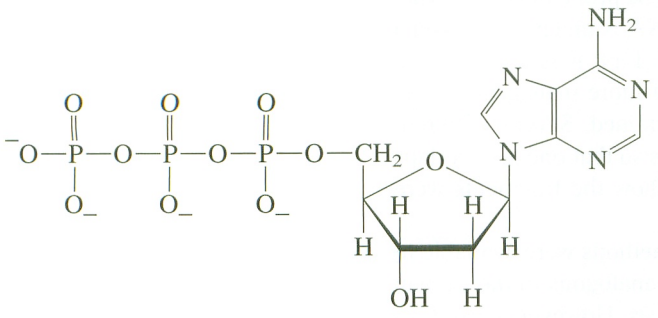
Cloning techniques have also played an important role in the development of DNA sequencing. Cloning allows many copies of the DNA to be made so that enough material to sequence is available.

Two methods have been developed to sequence the DNA once enough copies of the correct length have been obtained. The Maxam-Gilbert method uses chemical reactions to cleave the DNA at specific bases. The other procedure, called the chain-terminator method, was developed by Frederick Sanger. (Walter Gilbert and Sanger shared the 1980 Nobel Prize in chemistry for developing these sequencing methods. This was Sanger's second Nobel Prize, making him one of the few to have won twice. Recall that he also was awarded one for developing methods to sequence peptides.) The chain-terminator method is used for most sequencing today, so let's examine how this method works.

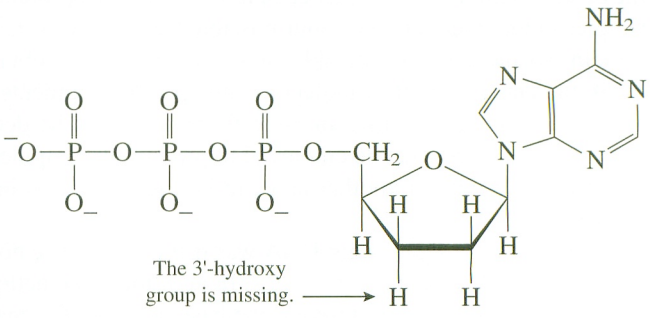
The chain-terminator method relies on the use of the enzyme DNA polymerase I, which catalyzes the synthesis of DNA. This enzyme makes complementary copies of single-stranded DNA, starting from the 5′ end and proceeding to the 3′ end. First the DNA to be sequenced, termed the template strand, is isolated. To begin the synthesis, the enzyme requires the presence of a small piece of DNA, called a primer, at the 5′ end of the chain to be synthesized. Because the template DNA is prepared by a restriction endonuclease, a few bases at its 3′ end are known. A primer that is complementary to this sequence is prepared by chemical synthesis. The enzyme then adds the bases that are complementary to the template DNA to the 3′-hydroxy group of the primer.

Sequencing is accomplished by incubating the template DNA, the polymerase enzyme, the appropriate primer and the four nucleotides, which are supplied as triphosphates. One of the nucleoside triphosphates is provided in radioactive form, so the

newly synthesized DNA is radioactive. Of critical importance, a small amount of the 2′,3′-dideoxynucleoside 5′-triphosphate corresponding to one of the nucleotides is also added. The structures of 2′-deoxyadenosine 5′-triphosphate (dATP) and 2′,3′-dideoxyadenosine 5′-triphosphate (ddATP) are as follows:



2'-Deoxyadenosine 5'-triphosphate (dATP)



2',3'-Dideoxyadenosine 5'-triphosphate (ddATP)

Because there is only a small amount of dideoxynucleotide present, it is incorporated only occasionally. However, wherever it is incorporated into the growing piece of DNA, the polymerization is terminated because there is no 3′-hydroxy group to which the next nucleotide can be attached. This results in a series of DNA fragments of different lengths, each terminated at the base that was added as its dideoxy derivative. The synthesis of DNA fragments using dideoxycytosine triphosphate is outlined in Figure 27.5.

To determine the sequence of the template DNA, four versions of the polymerization are run, each with a different dideoxynucleoside triphosphate. The reactions are analyzed using gel electrophoresis. In this technique the mixtures of synthesized fragments from each reaction are placed in separate lanes at the top of a layer of a polyacrylamide gel and a voltage is applied between the top and the bottom. Smaller molecules move faster than larger molecules, so they appear closer to the bottom of the gel

CTAGGCTTAGAC
GA
❶
CTAGGCTTAGAC
GA
dATP dCTP
dGTP dTTP
ddCTP
❷
GATC
GATCC
GATCCGAATC
Template strand
The DNA to be synthesized (the template strand) was produced using the restriction endonuclease AluI, so the first two bases are C and T.
❶ The primer that is complementary to CT is added, and hydrogen bonds to the 3' end of the template.
❷ The four nucleoside triphosphates, DNA polymerase, and a small amount of dideoxycytosine triphosphate are added.
The fragments that are synthesized all terminate with a dideoxycytosine residue at their 3' end. Remember, the fragments are complementary to the original sequence.

Figure 27.5

SYNTHESIS OF DNA FRAGMENTS USING DIDEOXYCYTOSINE TRIPHOSPHATE IN THE CHAIN-TERMINATOR SEQUENCING METHOD.

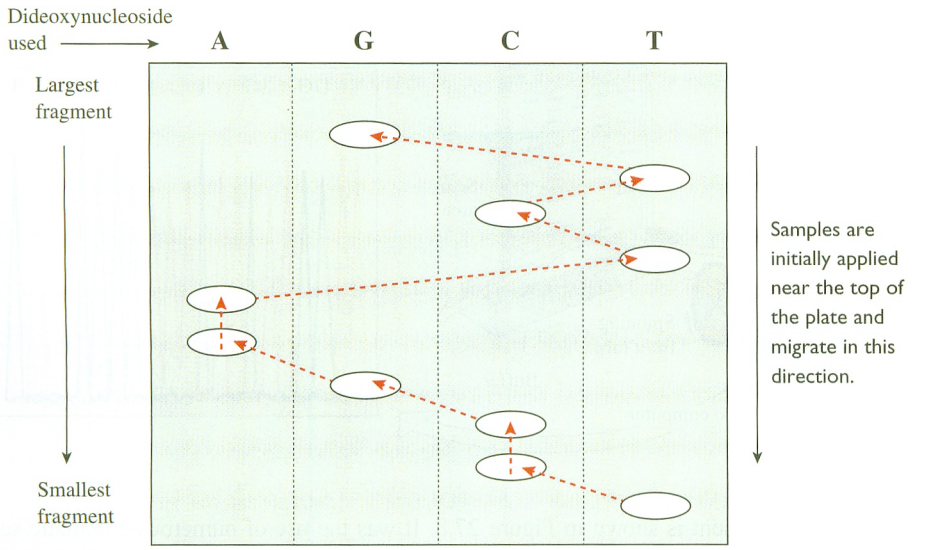


Figure 27.6

PATTERN OF GEL ELECTROPHORESIS SPOTS OBTAINED FROM SEQUENCING CTAGGCTTAGAC USING THE CHAIN-TERMINATOR SEQUENCING METHOD. The sequence that is complementary to the piece of DNA that is being sequenced is read starting at the bottom of the gel. The first two bases are from the primer (see Figure 27.5), so the smallest fragment terminates in T (the base complementary to A), so the lowest spot appears in the T lane. The sequence of the complementary strand is TCCGAATCTG.

when migration is halted. Once the separation is complete, a piece of film is placed over the gel. The location of the radioactive fragments can be determined by the spots that appear when the film is developed. Because the separation is based solely on size, the spot closest to the bottom of the plate is found in the lane of the reaction using the dideoxynucleoside base complementary to the first base in the template strand. The sequence of the complementary strand can be obtained by simply reading the gel from bottom to top. The electrophoresis gel pattern that would be obtained for the DNA fragment of Figure 27.5 is shown in Figure 27.6. A photograph of the autoradiogram of an actual sequencing gel is shown in Figure 27.7.

PROBLEM 27.9

Show the pattern of the gel electrophoresis spots that would be obtained from sequencing the DNA fragment CTTAGTTGCACCT using the chain-terminator method. The primer is GA.

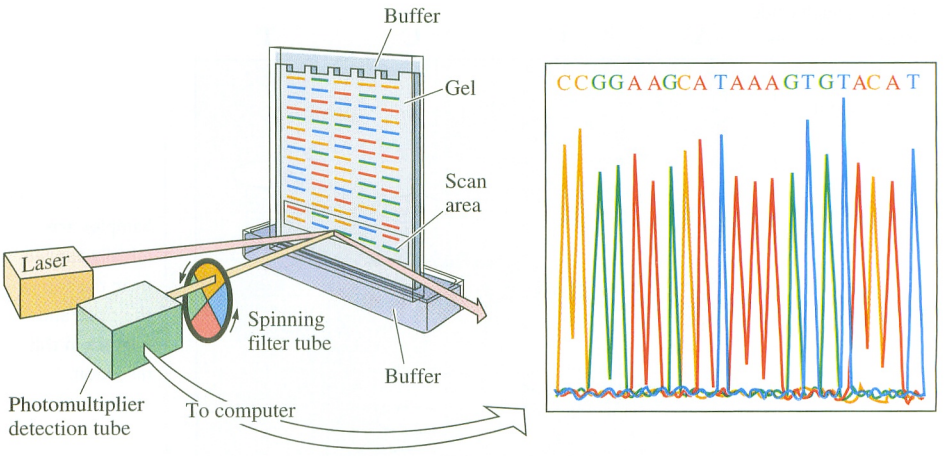
An automatic sequencing instrument has been developed that uses the chain-terminator method. To avoid the use of radioactive labels, a different color fluorescent dye is attached to the primer in each of the four reactions used to synthesize the DNA fragments. The mixture of fragments from all four reactions is then analyzed using electrophoresis in a single lane. A fluorescent spot appears for each polynucleotide of increasing size. The 3′-terminal base for each spot can be determined by the color of the fluorescence. The detection system is computer controlled, and the acquisition of data is automated. A schematic representation

Figure 27.7

AUTORADIOGRAM OF A SEQUENCING GEL.

Figure 27.8

SCHEMATIC DIAGRAM OF AN AUTOMATIC DNA SEQUENCER. Each lane represents a separate sequencing experiment.



of such an instrument is shown in Figure 27.8. It was the use of numerous automatic sequencers of this type that enabled the determination of the sequence of the entire human genome (2.9 billion base pairs) to be completed in 2001.

27.5 LABORATORY SYNTHESIS OF DNA

The ability to chemically synthesize DNA is quite important. This synthetic DNA is used in cloning, in site-directed mutagenesis (which enables the preparation of a protein with a different amino acid at a single site), and in the diagnosis and prenatal detection of genetic diseases.

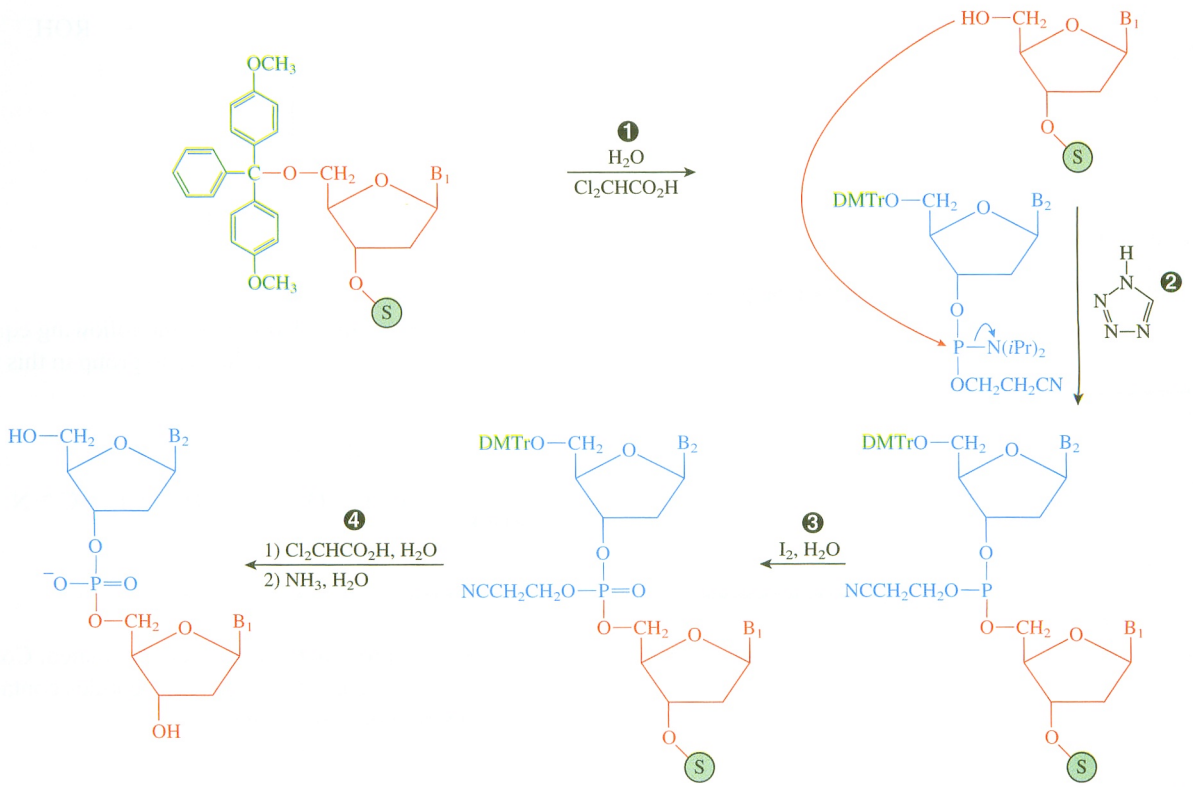
The basic strategy of polynucleotide synthesis is similar to that of polypeptide synthesis. Functionality in the nucleotide at which reaction is not desired must be protected. Then the nucleotides must be coupled. Next, the protecting group must be removed so that another cycle of coupling can be accomplished. After the desired cycles of coupling have been performed, any remaining protecting groups must be removed.

Currently, the most widely used synthetic method for DNA is the **phosphoramidite method,** a solid phase procedure that has many similarities to solid phase peptide synthesis. An example of the synthesis of a dinucleotide by this procedure is shown in Figure 27.9. First a protected nucleoside is attached to the surface of a silica particle through its 3′-hydroxy group. The free NH_2 groups of adenine, cytosine, and guanine are protected as amides. The 5′-hydroxy group is protected as a dimethoxytrityl ether. After the nucleoside has been attached to the silica, the trityl protecting group is removed by hydrolysis under S_N1 conditions. This is followed by reaction with a 3′-phosphoramidite derivative of the next nucleoside. The phosphorus in this compound is trivalent and is bonded to an oxygen of deoxyribose and to a diisopropylamino group. The remaining oxygen is protected as a β-cyanoethyl ether. The liberated 5′-hydroxy group displaces the diisopropylamino group, forming a phosphite triester with the desired bond to the phosphorus. Oxidation of the phosphite triester to the phosphotriester with iodine completes the addition of one nucleoside. This cycle of deprotection, coupling, and oxidation is repeated for each nucleoside that is to be added. After the desired number of cycles, the polynucleotide is removed from the silica particle by treatment with aqueous ammonia. This also removes

The first nucleoside is attached to the surface of a silica particle via its 3'-hydroxy group. This also serves to protect the 3'-hydroxy group during subsequent transformations. The NH_2 groups of adenine and cytosine are protected as benzamides, and the NH_2 group of guanine is protected as an isobutanamide. The 5'-hydroxy group is protected as a dimethoxytrityl ether (DMTr).

❶ Aqueous dichloroacetic acid is used to remove the dimethoxytrityl group in an S_N1 reaction.

❷ The unprotected 5'-hydroxy group is then reacted with a 3'-phosphoramidite derivative of the next nucleoside in the presence of tetrazole, which acts as a weak acid catalyst. (These phosphoramidite derivatives are now commercially available.) The diisopropylamino group is displaced by the 5'-hydroxy group, and the phosphorus–oxygen bond is formed.



❸ Iodine is used to oxidize the phosphite triester [$P(OR)_3$] to the phosphotriester [$O{=}P(OR)_3$].

❹ Dichloroacetic acid is used to remove the DMTr protecting group on the 5'-hydroxy group of the second nucleotide. At this point, a third nucleoside can be added and the cycle can be repeated. When all of the desired nucleosides have been added, reaction with aqueous ammonia is employed to remove the polynucleotide from the silica, remove all of the amide protecting groups from the bases, and to remove the β-cyanoethyl protecting group.

Figure 27.9

SOLID PHASE SYNTHESIS OF A POLYNUCLEOTIDE BY THE PHOSPHORAMIDITE METHOD.

the amide protecting groups on the bases and the β-cyanoethyl protecting group. The polynucleotide is then purified by gel electrophoresis or by chromatography.

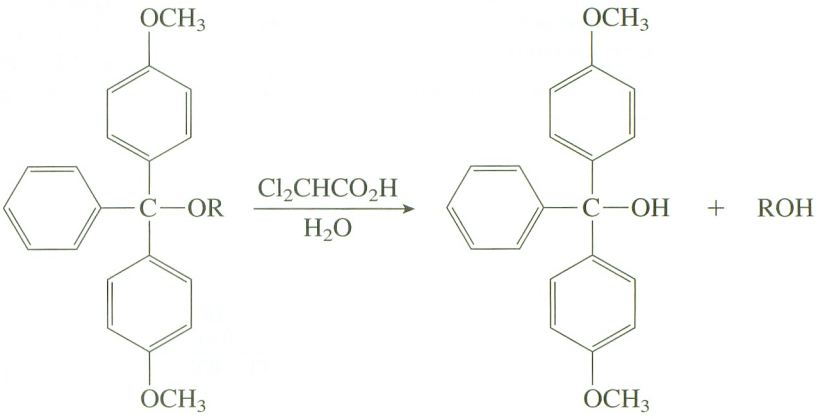
PROBLEM 27.10

Show a synthesis of this trinucleotide using the phosphoramidite method:

5′ end A-T-G 3′ end

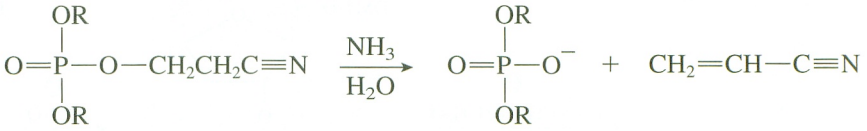
PROBLEM 27.11

Show the mechanism for the cleavage of the dimethoxytrityl ether. Explain why the methoxy groups accelerate the reaction.



PROBLEM 27.12

Removal of the β-cyanoethyl protecting group occurs according to the following equation. Show a mechanism for this reaction. What is the role of the cyano group in this reaction?



Like the solid phase peptide synthesis, this process has also been automated. Commercial automated synthesizers are available that can prepare polynucleotides containing more than 150 bases with a cycle time of about 10 minutes per base.

ORGANIC Chemistry Now™
Click *Mastery Goal Quiz* to test how well you have met these goals.

Review of Mastery Goals

After completing this chapter, you should be able to:

- Show the general structures of nucleosides, nucleotides, DNA, and RNA. (Problems 27.13 and 27.14)
- Show the hydrogen bonding that occurs between adenine and thymine or uracil and between guanine and cytosine. (Problems 27.19, 27.24, and 27.25)
- Understand the general features of replication, transcription, and translation.
- Understand the chain-terminator method for determining the sequence of DNA.
- Show a reaction scheme for the synthesis of a polynucleotide. (Problem 27.20)

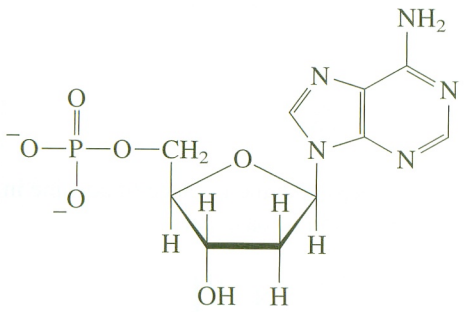
Additional Problems

ORGANIC Chemistry Now™
Assess your understanding of this chapter's topics with additional quizzing and conceptual-based problems at http://now.brookscole.com/hornback2

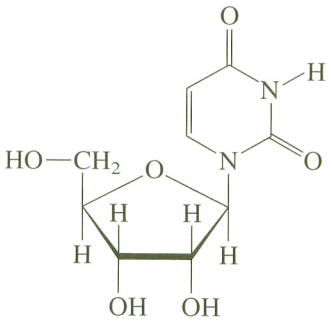
27.13 Draw the structures of these compounds:
- **a)** Adenosine
- **b)** ddCTP
- **c)** Deoxythymidine 5′-monophosphate

27.14 Provide names for these compounds:

a)



b)



27.15 Explain why cytosine is a much stronger base than thymine or uracil.

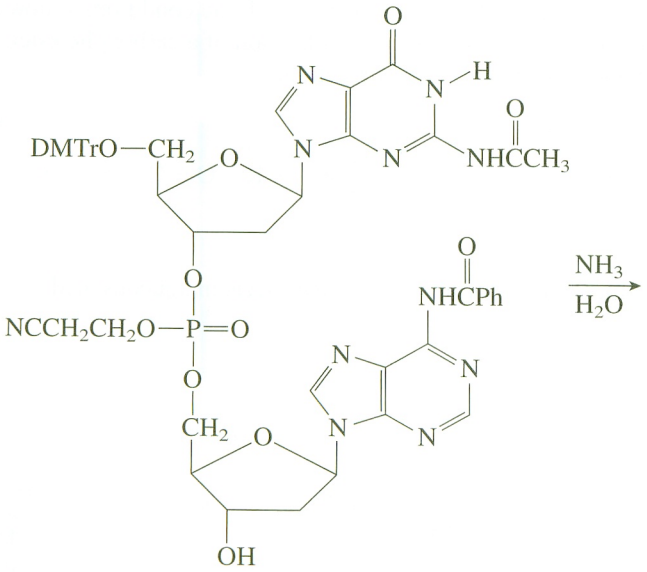
27.16 Show the tautomers of uracil.

27.17 Show the sequence of the DNA and of the RNA that is complementary to this piece of DNA:

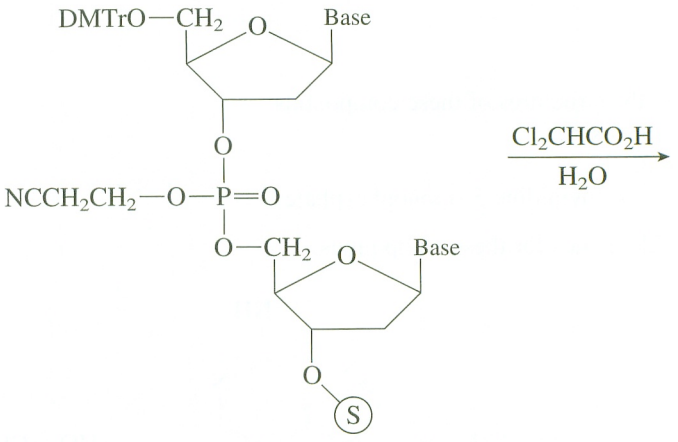
5′ end GCTTATGC 3′ end

27.18 Show the products of these reactions:

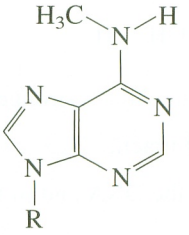
a)



b)



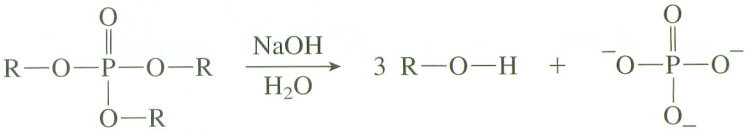
27.19 Some organisms use this methylated adenine in place of adenine in their DNA. Explain how this affects the formation of base pairs.



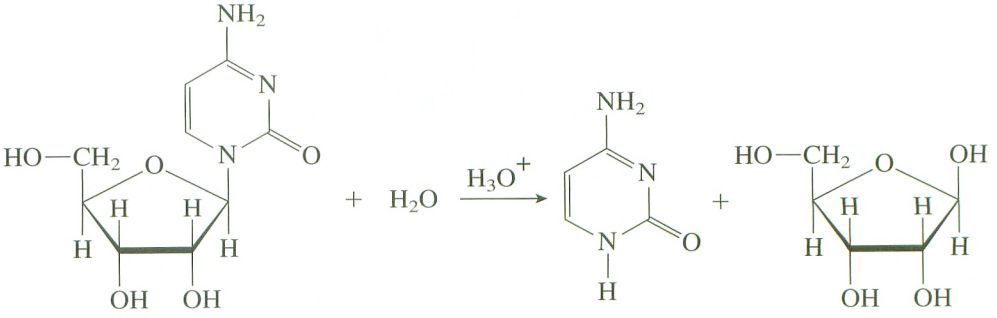
27.20 Show a synthesis of this trinucleotide by the phosphoramidite method:

5′ end GCT 3′ end

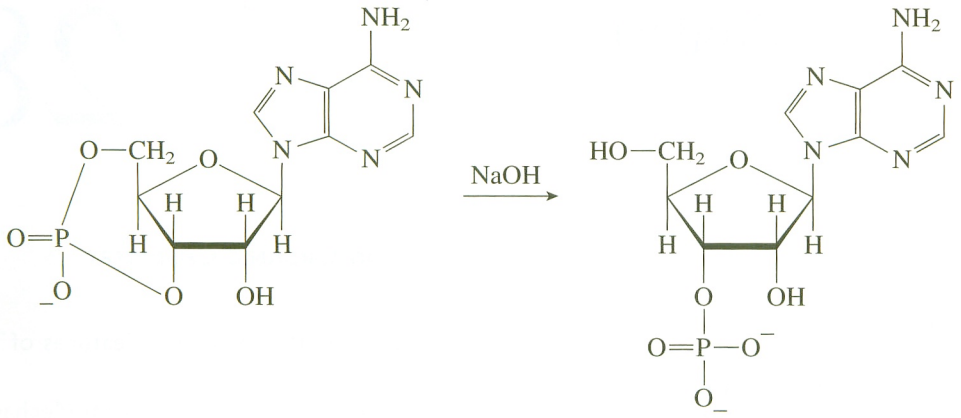
27.21 The hydrolysis of a phosphate ester under basic conditions follows a mechanism that is similar to that for saponification of a carboxylic ester. Show the steps in the mechanism for this reaction:



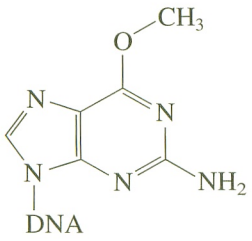
27.22 Show a mechanism for nucleoside hydrolysis in aqueous acid.



27.23 Explain why this cyclic phosphodiester gives mainly the 3′-monophosphate rather than the 5′-monophosphate when it is treated with aqueous sodium hydroxide:

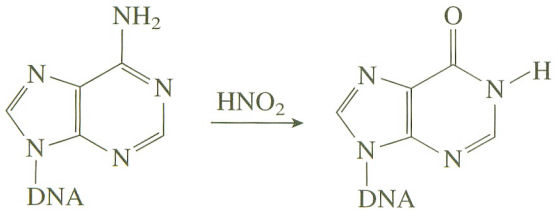


27.24 Some of the nucleophilic atoms in the bases of DNA can be alkylated in an S_N2 reaction with highly reactive electrophiles. For example, the oxygen of guanine can be methylated to form *O*-methylated guanine. The presence of an *O*-methylated guanine often results in mutations caused by the incorporation of thymine rather than cytosine during replication. Show how base pairing explains this observation.



O-Methylated guanine

27.25 Nitrous acid (HNO_2) is a mutagen (a compound that causes mutations). It is thought to operate by changing the structure of some bases in DNA. For example, it is capable of converting adenine to hypoxanthine. Explain how this might cause a mutation.



Adenine Hypoxanthine

Do you need a live tutor for homework problems? Access vMentor at Organic ChemistryNow at **http://now.brookscole.com/hornback2** for one-on-one tutoring from a chemistry expert.